DOCUMENT RESUME

ED 376 187 TM 022 205

AUTHOR Broadfoot, Patricia
TITLE Approaches to Quality Assurance and Control in Other Countries.
PUB DATE Apr 94
NOTE 18p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 4-8, 1994).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Accountability; *Centralization; College Entrance Examinations; *Educational Assessment; Elementary Secondary Education; Foreign Countries; Higher Education; Models; Norm Referenced Tests; *Quality Control; *Social Influences; *Teacher Responsibility; Teacher Role; *Testing Problems
IDENTIFIERS Australia; External Examination Program; France; Germany; New Zealand; Sweden

ABSTRACT

The educational assessment quality assurance and control practices of Germany, Sweden, France, New Zealand, and Australia are compared to illustrate the considerable variety of approaches and some of the social factors that contribute to this variety. In Germany there is considerable variation from one province to another, and quality assurance and control in assessment are largely part of the professional responsibility of teachers, based on the long-standing legitimacy of the procedures in use. In Sweden there is also considerable emphasis on quality assurance through developing the expertise of teachers regarding assessment and then trusting them to carry it out. In France there is a heavier reliance on externally set examinations and central control. Education in New Zealand has been dominated by external assessment. A new quality-assurance framework is being implemented to move the system from a norm-referenced examination system to one that assesses achievement against recognized standards. Every Australian state has its own system, and the formal provisions for quality control and assurance are focused on secondary school and entrance to higher education. Three figures illustrate the discussion. (Contains 25 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 376 187

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN **OTHER THAN PAPER COPY** HAS BEEN GRANTED BY

PATRICIA BROADFOOT

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)"

UNIVERSITY OF BRISTOL

APPROACHES TO QUALITY ASSURANCE AND CONTROL IN OTHER COUNTRIES

by

Patricia Broadfoot

Paper presented as part of the Symposium:

Enhancing Quality in Student Assessment

by

The British Educational Research Association

at

The American Educational Research Association Conference

New Orleans, April 1994

TM022205

This paper has been prepared for the annual conference of the American Educational Research Association, New Orleans, April 1994

The symposium: Enhancing Quality in Student Assessment

Title: **APPROACHES TO QUALITY ASSURANCE AND CONTROL IN OTHER COUNTRIES**

Author: Patricia Broadfoot

Presenter: Patricia Broadfoot

Correspondence to: Professor Patricia Broadfoot, University of Bristol School of Education, 35 Berkeley Square, BRISTOL BS8 1JA, U.K.

Telephone: U.K. (0272) 303030 Fax: U.K. (0272) 251537

APPROACHES TO QUALITY ASSURANCE AND CONTROL IN OTHER COUNTRIES

Introduction

One of the great strengths of comparative educational research is that it can document contrasts between countries in the way that they seek to achieve similar ends. In so doing, it not only helps to illuminate complex relationships of cause and effect but it can also stimulate the questioning of existing practice (Crossley and Broadfoot, 1992). At a time when countries are actively seeking to 'borrow' policies and practices from each other (Phillips, 1989), the provision of comparative information in an analytical and systematic way is potentially an important avenue through which educational research can make an impact and thus contribute to the improvement of practice. Comparative studies also underline the equally important caution that all such differences are reflections of the culture that produced them and thus can never be simply translated from one cultural setting to another.

Mindful of both the strengths and the dangers of international comparisons, this paper adopts such an approach in order to contrast different national approaches to enhancing assessment quality. From the various national case-studies briefly presented here, a picture of a considerable variety of practice emerges. Typically, these practices reflect different cultural and institutional traditions and different assessment priorities. Typically, too, they are in a state of transition as new pressures on the education system are leading to the development of new approaches to assessment. In many countries, the need to promote 'higher-level' skills, skills that encompass both cognitive and affective domains; the need to assess individuals as part of a team and their capacity to be self-critical as well as the need to assess a great variety of more vocational competences within the mainstream is prompting a movement away from standardised tests and written examinations to more 'on the job' assessments in classrooms or work places. Baker, O'Neil and Linn (1991) summarise the attributes of the alternative assessment approaches now being developed on a widespread basis in the US as including: the use of open-ended tasks, a focus on higher order or complex skills; the employment of context-sensitive strategies; performance-based sometimes over an extended time period; involving either individual or group performance and possibly involving a considerable degree of student choice' (p.38).

In a country such as the United States which has long been dominated by psychometric approaches to assessment, the proposed introduction of a national system of examinations under the New Standards Project which would use this kind of approach is raising quite novel questions concerning assessment quality. Assessment quality has been both defined and assured in the past by the empirical development of test items, standardised and secure administration and scoring practices, and the norming of tests on carefully selected populations.

Where the nature of what is to be assessed requires a more open-ended longer-term or teacher-controlled assessment a more flexible way must be found of standardising the resultant assessment information. In our terms, some of the options concern quality 'assurance' devices including the defining of standards and assessment criteria, the accreditation of institutions, visits by verifiers (see Fig.1) or agreement panels. Alternatively, the emphasis may be put on quality control (see Fig.2) by means of a scaling test, inspections of work samples or a panel review - that is a means of standardising the assessment information generated. In each case the aim is to find a technique which will be as valid as possible whilst ensuring the assessment is sufficiently rigorous to justify public confidence in decisions made on the basis of it.

If there is no optimum way of achieving this balance there is also no common path towards it. As a technique of power which controls to a considerable extent the distribtuion of life chances, assessment is inevitably a highly political activity. Decisions concerning changes in practice are therefore likely to be as much a reflection of prejudice and perceived interests as they are the result of technical arguments concerning quality. Thus apart from producing a 'map' of possibilities, the identification of aspects of particular national contexts which lie behind the commitment to certain assessment traditions and practices is in turn the basis for generating a more general understanding of the key variables which need to be taken into account in any attempt to change such practice. With these goals in mind, this paper reviews the quality assurance and quality control practices of five countries: Germany, Sweden, France, New Zealand and Australia. These examples have been chosen in order to illustrate the considerable

variety of practice which exists internationally and briefly to identify some of the key social factors which have led to these differences of approach.

Madaus and Kellaghan (1991) suggest that established European practice may well have an influence on US assessment practice. However, Madaus and Kellaghan distinguish the long-standing emphasis of European examinations on content with the associated assumption that examinations will heavily constrin content from the current US preoccupation with promoting skills and suggest that this is likely to sustain continuing differences in approach. Madaus and Kellaghan further distinguish the different assessment traditions of Europe and the US in terms of the level of public confidence in teachers who in Europe are trusted to determine and grade examination questions. Because of this they suggest in Europe, far from being separated from it, assessment is integrally related to the curriculum. However, within this global distinction between the very different traditions of the US and Europe (and, by implication, Australia and New Zealand) there are some important differences of approach among the countries that have been influenced by European quality assurance and control traditions which are worthy of further exploration.

Quality Assurance and Control in Germany

The defining context of German education, especially in the former West Germany, is the right of the individual 'Länder' to make and implement its own educational policy. Hence there is considerable variation from one province to another. There is now, however, agreement between the 'Länder' over broad common conditions for the structure, conduct and subject composition of the key educational qualification - the 18+ 'Abitur' and major qualifications are typically accepted as equivalent between the 'Länder'.

The other defining characteristic of German educational provision is its strongly centralised character. In most 'Länder' the curriculum is centrally determined according to specific statements of what is to be learned at each age level. Partly because of the very strongly centralised control of content there has been little national concern over standards, and the assessment procedures remain very largely the responsibility of teachers even for certification and selection purposes with minimal external intervention. This is partly because the use of a six-point mark scale (Notenskala), ranging from 1 'very good' to 6 'irretrievable failure' is a well known feature of German education and is well understood by all associated with German schooling. Even the important decision concerning secondary school destination is made on the basis of continuous assessment by teachers. This approach to assessment stresses the key role of the teacher as a 'professional assessor' such that no further moderation or control mechanism is needed (although parents do have the right of appeal).

Given its influence, it is surprising that the legitimacy of the traditional marking system in German education is not underpinned by any mechanism for acquainting teachers with standards in other schools. There is now some evidence of concern within the system in this respect in recognition that teachers may well not have a good grasp of other schools' standards and that standards may vary between one Länder and another.

Access to upper secondary education and higher levels of vocational education depends on the acquisition of an appropriate leaving certificate. The judgement is typically made on the basis of assessments made within the school including systematic half-yearly reports and discussions among the teachers of a class. These reports are based on formal internal school tests. Whilst these tests are carefully constructed and pupils are warned of the time and content of them in advance, they are within the control of individual teachers.

Assessment for the university-entrance 'Abitur' score is based on a combination of examinations and continuous assessment during years 12 and 13. Where an exam mark deviates significantly from coursework assessment, the student must take an oral in that subject, or he/she can request an oral in the hope of marginally raising an exam mark. The total 'Abitur' score is calculated by combining results gained in various courses during the last two years of schooling with results of the final exams.

Although there are some in which the questions are centrally set and presented to schools as an unseen examination, in most of the 'Länder' the 'Abitur' questions are set by the candidate's own teacher. Teachers are provided with national and 'Länder'-produced guidance to help them in the generation of

test questions. Typically the exam papers are marked by the candidate's own teacher and then second marked by a colleague from the same department. The school is responsible for quality control in checking that the administration of the Abitur has been in strict accortdance with national requirements laid down by the KMK.

Thus, except for some external checking of samples of scripts, quality assurance and quality control in Germany is very largely part of the professional responsibility of teachers..

In their report published in 1987 entitled *Education in the Federal Republic of Germany: Aspects of Curriculum and Assessment*, HMI record how cheap to run the German assessment model is without the elaborate paraphernalia of external involvement typical of some other countries - such as the U.K. - where teacher assessment is used - a point also made by Sadler (1991) concerning the internal assessment used in Queensland, Australia, which he estimates saves two thirds of the cost of an external assessment system. The relatively low key German approach to comparability may be because the high stakes point of the system, namely the 'Abitur', was traditionally a qualifying exam for university with all successful students being entitled to enrol. As increasing numbers of university departments have introduced a *numerus clausus*, concerns about the comparability of the assessments made have become more overt.

In relation to the other 'high stakes' point of the system, namely the eleven-plus-type selection, again this was not a major problem in the past as many German parents did not aspire for their children to go to the gymnasium. The increase in parental aspirations which might have caused a problem has been matched by a softening in the allocation procedures as described above, so that the elevation of the assessment into an excessively high stakes mode has been avoided.

Thus quality assurance in Germany is essentially a professional matter. The structural purposes and long-standing legitimacy of the procedures in use, coupled with the high traditional status of teachers in Germany, combine to make unnecessary any more formal provision for comparability. Furthermore, the pervasiveness of the notenskala system means that pupils are under great pressure to achieve demonstrably and continuously. Equally significant is the process of internal socialisation in the primary school years whereby children gradually learn to see grading as personal affirmation (Chisholm, 1987).

Quality Assurance and Control in Sweden

Like Germany, Swedish education provision has traditionally been characterised by a high degree of central control at both school and higher education levels. Although a margin of decentralisation has recently been introduced there is still a centralised curriculum. However, the approach to moderation in Sweden is very different emphasising a bureaucratic, statistical moderation approach. This is perhaps because of the well-known Swedish preoccupation with equality and hence, as in the United States, with the need for transparent reliability.

At the end of each three year stage of the nine-year compulsory school which provides comprehensive education for all children from age seven to sixteen, teachers are provided with nationally produced but voluntary diagnostic tests. The aim of these tests is to identify whether students have reached the minimum competency required in the core subjects of Swedish and maths at the end of each stage. Similar tests are also used in the upper secondary vocational lines. The tests are designed to provide teachers with information so that they can compare their pupils' results against the national average. One of the effects of the tests is to focus teachers on the essential competences to be acquired, and the tests are deliberately designed to do this. Where inadequacies are identified in this way, teachers are expected to provide remedial support.

Entry to the chosen course in upper secondary school is now not subject to any formal hurdle and students are assessed at the end of the basic school only on the traditional five point marking scale. Comparability of marking between schools is assured by the use of centrally-generated tests, the results of which are used to calibrate teachers' own marks. The nationally normed test results received from the schools are analysed and (some revised) scales determined in order that teachers can compare their own pupils' achievement with those of other teachers throughout the country and use this information in giving marks to individual pupils within their classes.

A similar test for upper secondary school is compulsory for all the main subjects. The tests are published each year and therefore annually produced. Teachers adjust the marks students achieve on their own tests, set in each subject during the year in order to correspond with the results achieved by their classes in the national scaling tests. These tests, which are open to public scrutiny and which have a clear effect on the grades which students can achieve on leaving upper secondary school, are 'high-stakes' and exert a not inconsiderable wash-back effect on the curriculum. This is taken into account when the tests are devised and considerable efforts are made to make sure the tests reinforce the whole range of skills embodied in the curriculum.

Since 1991 all students have been able to opt for selection for higher education based not on moderated school marks but on scores achieved on an alternative test which, like the American Scholastic Aptitude Test which it resembles, is a multiple choice test. This was originally provided for adult entry to higher education but is now taken by some 40% of school students. Marks are converted to standardised scores by means of statistical moderation.

Thus in Sweden, as in Germany, there is very considerable emphasis on quality assurance in assessment through developing teachers' expertise in assessment and then trusting them to carry it out. Unlike Germany, however, it is more of a 'belt and braces' approach in that quality control procedures in the form of a scaling test are also involved - most particularly at the key decision points in a student's career. Not surprisingly, this approach causes considerable unhappiness among teachers forced to change their marks and can seem arbitrary in use - a problem that the Government has sought to address in part at least by providing more guidance for teachers concerning the assessment process, such as that on marks and marking in the upper secondary school, published by the National Board of Education in 1990. It also reinforces a norm referenced approach to assessment. It is clearly not universally popular with students either as increasing numbers opt for the alternative multiple choice test.

Quality Assurance and Control in France

France has many similarities with Germany in the association of the traditionally high-status upper secondary teachers with a strong emphasis on professional responsibility in assessment. However, the tensions caused by the innovation of continuous assessment alongside the more familiar external examinations, illustrate the heavier reliance that has traditionally been placed on externally set examinations in France. Indeed, Kreeft (1990) characterises France as a classic example of an external examining system. For example, the first formal assessment - the Diplôme National du Brevet which is taken at the end of collège before proceeding to the lycée, is based on the combination of an external exam in French, maths, history and geography and on the results of continuous assessment. It is something of an anomaly, because, whilst the locally set examination is rigorously standardised, there is little or no moderation of the continuous assessment awarded by the school which tends to be over-generous in its marking. Even so, in many places, the percentage of students achieving average marks is too low so that the whole distribution is moderated up to ensure 20% of candidates achieve an acceptable standard. This results in inequalities between different departments and penalisation of the more rigorous, leading to widespread doubts about the value of the results (Caroff, 1991).

It is not without significance that despite repeated calls to reform the prestigious 18+ baccalauréat examination into a more school-based exam, not least because of the enormous cost of teacher time in externally examining what is now a huge number of students, students, parents and teachers continue to give very strong support to the notion of an anonymous external examination as a guarantee of equality against the influence of teachers' values and schools of different status. Indeed, although largely a symbolic role, the Baccalauréate's Napoleonic origins and its domination of the education system for nearly two centuries means it has almost become a defining symbol of the integrity of French culture. Yet, despite this rhetoric it is widely accepted that there are variations between different subjects and different geographical areas in current procedures. These variations have become more acute since 1966 when the responsibility for the exam was entirely devolved to the 'recteur d'academie'. Although the 1983 Prost Report on the reform of the lycée argues for a considerably increased measure of continuous assessment, little progress has been made in this respect. In 1991 Lionel Jospin again tried to introduce continuous assessment in the baccalauréate but without much success.

Those who have been involved in continuous assessment - for example with the more vocational Baccalauréate Professionel - appear to doubt the national value of a qualification in which teachers have a say in the results of their own students. This may well be both a cause and an effect of the fact that no non-traditional moderation procedures appear to have developed to support this kind of assessment. It also marks a stark contrast to Germany where this would be regarded as normal.

Thus with its traditionally high measure of central control of the system, the issue of comparability in France has not been as significant as in some other countries. As with the 'Abitur' in Germany, the traditional right of access to higher education to all holders of the baccalauréat may have played some influence in this, although the advent of a *numerus clausus* in popular faculties is certainly raising the stakes in some 'baccalauréat filières'. There is a very considerable wastage after the first year of university and the 'propédétique' exam which also helps to lower the stakes of assessment during the course of schooling. In France, the notion that all schools are equal and the resolute refusal, *de jure* at least until recently, to recognise that schools might differ from each other in terms of the standards produced, has also contributed to reducing the significance of moderation as a formal issue. Undoubtedly the establishment of the baccalauréat exam for nearly two centuries since its foundation in 1808 contributes enormously to its legitimacy. This is both a strength - in that it tends to reduce potential pressure for apparently more effective quality control procedures - and a weakness - in that it is very hard to bring about any kind of change.

The two final examples in this paper provide further contrasts to the American and European traditions in describing two countries - New Zealand and Australia where there is a strong influence of British traditions coupled with a range of idiosyncratic local developments. These have combined to produce some rather different approaches to quality assurance and control.

Quality Assurance and Control in New Zealand

Perhaps because of the strong British influence, New Zealand education has long been dominated by external assessments. Indeed, until recently pupils were subject to an exam in each of the final three years of secondary schooling: the School Certificate taken at the end of Year 10, sixth form certificate at Year 11 and bursary at Year 12. School Certificate has traditionally been a school-based examination moderated by an externally devised scaling test, not unsimilar to the Swedish model. This test has been the bone of much contention, not only because of its arbitrary impact on school certificate but because it is used to constrain schools in their allocation of marks in the following year. This scaling test was abolished in 1992 and indeed the whole system of formal qualifications has now been overtaken by a major national initiative - the National Qualifications Framework - designed to move the system from a norm-referenced examination system to a system which assesses achievement against recognised standards. This initiative involves a change to unit based learning and to modular assessment. The National Certificate units will cover a spectrum through general subjects to clearly vocationally-oriented subjects. Existing school subjects will be transformed into general units.

Given that all qualifications in the post compulsory stages of education although not, at present, degree and post degree level, will be in a common framework and system of organisation, this will represent a radical change in the basis of quality assurance and control in the awarding of qualifications. The New Zealand Qualifications Authority (NZQA) has taken the decision to devolve equality management progressively to providers as an integral part of their delivery of services, with quality audits by the authority (NZQA, 1991).

It is envisaged that quality assurance will be based on registration, a process which ensures that private training establishments meet fundamental standards of delivery which safeguard their clients; approval, the formal recognition of units of learning and groups of units which make up qualifications by the authority; and accreditation, the formal recognition that a provider is capable of delivering units to the standard specified. The responsibility for the quality of education and training lies first and foremost with providers in partnership with their clients and the role of the qualifications authority is envisaged as one of ensuring that providers have quality control and assurance processes in place. This requires quality audit, the independent checking of the quality processes of providers. Quite what all this will mean in practice has not yet been worked out in detail (Q.A. News Issue 10, 1991) but two documents issued in 1992 by NZQA illustrate the high profile of moderation issues in the development of the new system and

the possibility that there will be devolution of the responsibility of providing for quality assurance and control.

The system already requires that registered units have clearly expressed performance criteria. Accredited providers must have quality management systems including policies and procedures that cover teaching and administration as well as assessment and reporting and provide for on-going evaluation of the quality of such procedures. They will be subject to external audit.

National moderation can be provided by NZQA itself, a national professional or industrial association, providers or private consultants. Alternatively moderation can be locally conducted on the basis of criteria laid down by a national standards body if they can confirm that local arrangements meet the criteria. Twelve possible approaches to moderation are offered including: the use of exemplars and benchmark materials; recognition of 'expert' assessors; external written examinations; statistical moderation; common assessment tasks and reference tests; item banks; distance moderators; external assessors (to assess locally); external moderators with site visits (to confirm local assessments); external moderators with panel meetings (moderator in charge); consensus panels (group decisions); and panel networks (NZQA (1992).

Thus it would seem that New Zealand is in the process of moving from a moderation system based on closed statistical procedures to one based partly on criterion-referenced assessment coupled with quality assurance procedures operated within the learning institutions themselves. This will be combined with a professional model involving a reliance on the increased professionalism and expertise of teachers in the administration of national assessment. In this respect, New Zealand has strong parallels with France and to an extent with Scotland. It remains to be seen, however, how significant these various initiatives will be in breaking down the traditional dominance within the educational system of the very high stakes, Bursary and Scholarship examinations, with their reliance on traditional formal external examining procedures and traditional forms of quality control.

Quality Assurance and Control in Australia

Each state in Australia has its own education system and only two are included here to illustrate their contrasting system of moderation. As a general rule formal provision for quality assurance and quality control in assessment is focussed on the upper secondary school stage and particularly entrance to higher education. Moderation in the two states to be discussed here is related to entrance to tertiary education and is carried out either because of the emphasis on school-based assessment (Queensland) where there is no external school leaving exam. or if there is an external certificating exam, statistical moderation is used to scale scores to provide a single aggregate for entrance to university. In some cases there is both group moderation and scaling to produce a ranking of students.

Queensland

There have been no external exams in Queensland for twenty years and assessment is therefore all based in school (Baumgart & McBride, 1992). The first formal assessment point is at year 10. Formerly the Junior Certificate, the school-issued certification and exit statement is now absorbed into a student portfolio which was introduced in 1989 for all year 10 pupils in State schools and 1990 for non-State schools. This portfolio shows the academic and personal achievements of the student and includes documentation concerning their abilities, skills and interests, which provides a comprehensive picture of the student's accomplishments, attributes and involvement. Given the essentially descriptive nature of the year 10 documentation, there is no move to provide formally for quality control of the data although efforts are being made on quality assurance to help teachers improve the quality of their assessment procedures.

At year 12 however, whilst the assessments are still school-based, there is considerably more attention given to external moderation in the critically important context of university entrance qualifications. School based assessment is on a five-point rating scale still moderated by panels of teachers who review the quality of work from different schools i.e. through consensus moderation. A separate procedure is used to scale the school-based achievement scores across subjects and across schools. This is done in order to calculate for each pupil an aggregate score and to compile an order of merit list for allocating

places in higher education. All students who wanted to go to tertiary education took a scaling test: the Australian Scholastic Aptitude Test (ASAT). This test was given to all students and used to scale scores awarded by teachers first within schools, and then across schools (this device of using a common test as a scaling device is widely used in Australia, and is further discussed below). The scaling is designed to allow their achievements in different subjects to be projected on to a common scale; with the aggregate scores for all students in a school yielding a distribution with the mean and standard deviation corresponding to those of the ASAT scores. This tertiary entrance score (TE) was convenient for universities because it allowed them: a very simple (some might say simplistic) form of selecting students without reference to additional information. From 1986 the school-based assessment has been on a criterion-referenced model with defined performance criteria linked to the five-point scale. This system of criterion-referenced assessment also involves group moderation.

The latest arrangement following the Viviani Review (1990) involves the use of school-based assessment with the single TE score replaced by a Student Education Profile containing a wider number of scales to be used in different combinations for selection into different university courses. This is significant as a move away from the single aggregate score so typical of higher education selection arrangements in Australia. Certification at the end of senior school thus has two levels: the senior Certificate records performance in the final two years of schooling as measured by school-based assessment. Reporting is again on a five-point scale together with results on a Core Skills Test. The Core Skills Test is new and it is given to all year 12 students. The Core Skills Test maps on to the curriculum and involves four separate tests: two multiple choice, one essay and one short response assessing cross-curricular skills; the core skills are also reported on a five-point scale. The second certificate, called the Tertiary Entrance Statement, provides a rank order for students for university selection. The students are given an overall position (OP) which is a single aggregate (as with the old TE) but on a more detailed 25 point scale derived by averaging, with equal weight, scaled scores on the best five subjects for each student (the school based assessment is scaled by using the core skills test result) but in addition each pupil is given up to five positions in separate fields. The field skills are: extended written expression; short written communication in English or a foreign language; basic numeracy; complex problem solving; substantial practical performance. Normally students will get a field position score in two or three of the five fields. The idea is that universities will be obliged to look at the students' overall profile of results rather than just relying on the single aggregate score.

At lower levels of schooling, assessment of students is the profesisonal responsibility of teachers and not subject to formal quality control. Thus we can see that in Queensland the context for moderation is that of tertiary entrance with the approach a mixture of statistical scaling and group moderation. The central concerns are the need to provide for comparability in a system which is heavily dependent on school-based assessment and questions over the validity of scaling to produce a single aggregate score. One anxiety about the new system in Queensland is that the moderated/scaled result is going to be fed back to the candidate. This has not happened before and there is some concern that it will lead to far greater emphasis being put on to the preparation for the core skills test.

New South Wales

In New South Wales students are assessed for the School Certificate at the end of year 10 and Higher School Certificate (HSC) in year 12. A State where certification procedures have traditionally been dominated by external examinations, the Board of Studies introduced a greater measure of school-based assessment in 1990 to provide students, parents and employers with more detail concerning achievement. External control of standards will be sustained through the continued use of an external reference test in English, Mathematics and Science. The results of this test are used to allocate a range of grades to each school. The school then awards these grades to students on the basis of their performance on the school's own assessment programme. In the three tested subjects, statewide distributions of (10%A, 20%B, 40%C, 20%D and 10%E) are applied. From 1991 all other subjects have been graded A to E though on the basis of performance descriptors issued to schools rather than on the results of an external refernce test. These performance descriptors, including course-specific descriptors, - elementary to excellent - will be printed with the grades on the year 10 record of achievement, for example:

A - A grade indicating excellent achievement in the course. The student has an extensive knowledge and understanding of the course content and can readily apply this knowledge. In addition, the student has achieved a high level of competence in the processes and skills of the course and can apply these skills to new situations. (Australian Cooperative Assessment Program Newsletter, No.2, 1990)

The intention has been to encourage teachers to apply criterion, rather than norm-referenced standards - an aim which will be helped by there being no requirement for a fixed state-wide normal distribution.

In year 12, however, the results on the HSC are scaled to produce a tertiary entrance score (TES), an aggregate single score which is used for tertiary entrance as in the Queensland system. However students do not take a common reference test for scaling purposes, the scaling is done instead on an iterative system using the unbiased mean grade system or the UBMT (unbiased mean total). For example, if English is the subject under consideration, the UBMT for any candidate taking English is the sum of the grades obtained in all other subjects attempted, divided by the number of those subjects, or the mean grade in all other subjects attempted, i.e. minus English. This is therefore a strong version of moderation using statistical scaling.

The HSC involves an element of school-based assessment, commonly 50% of the marks. From 1986 the school-based mark and the examination mark were reported separately; since 1988 the certificate has also included a report of the combined marks in decile bands. The scaling process was also used to produce a mean and standard deviation for each group of students in the school-based assessment with the same mean and standard deviation as obtained by the group in the external exam. Originally this scaling involved a linear transformation and it was thought that "examiners and/or schools on occasion sought to modify the shape of the distribution to advantage some students" (Baumgart & McBryde, p.6). Thus the system has moved to one of area transformation of scores rather than linear transformation of scores.

Calculation of the TES has produced regular problems and the scaling of community languages has been particularly difficult. In 1987 the Board of Senior Secondary studies (the examination board) removed itself from the production of the TES and left it to the universities. Thus, currently, scaling across subjects has been abandoned and scores for each candidate for each course are reported raw without any attempt to adjust for the nature of the candidature. For one year all universities except the University of Sydney, used a TES formed without scaling across subjects, but currently all universities use a system to produce a TES based on scaled scores across subjects (the current situation is that a TES is calculated first, and then a tertiary entrance rank (TER) is reported in percentage form in steps of 0.05).

The TES has been based on the best ten units offered by the student unless particular departments or universities require some unit of the curriculum to be included. This has meant that in the past, for example in 1991, 94% of medical students and 45% of law students entered the University of Sydney from the HSC with four maths units out of their ten on the HSC, while less than half of these students offered English at all in their TES. The current proposal is that there be some specification about which ten units be used for the TES including at least one unit of English and one unit from each of other areas of the core curriculum; not more than six units from any one field of study. In short they are moving towards a core curriculum to be included for tertiary entrance.

Thus in New South Wales the key focus for quality assurance and control is that of tertiary entrance. The approach to providing this is almsot entirely that of statistical scaling and the issues are again to do with the validity of producing an aggregate score. Although there is an element of school-based assessment there is no group moderation process. Nevertheless,

> "while the statistical scaling procedures used for the New South Wales HSC do not allow for teachers to be involved in moderation, involvement in the marking procedures is reported by many Head Teachers (of subject departments) as a highly valuable professional development activity for teachers of senior students. The experience of marking examination papers of students from across the state provides teachers with an idea of appropriate 'standards' such that when these teachers work on setting school based assessment tasks the following year they have a good understanding of the requirements of the Board of Studies" (Nelson, 1993).

As in other States, New South Wales is currently concerned to introduce more extensive quality assurance procedures in schools. The Quality Assurance School reviews include a concern with the assessment and reporting practices in a school - a part of its aim is to help schools improve in this respect. Nevertheless, as an editorial in the *Sydney Morning Herald* (30.11.93) makes clear, New South Wales is firmly committed to external assessment as the key to quality: 'the intention behind the reforms ... was to create a fairer, more efficient and effective school system. In New South Wales this was done by empowering individual schools at the expense of the previously dominant central bureaucracy, the HSC examination, though was kept as the credential focus of the system' (p.12).

Thus, as Withers (1992) points out, Australian 'school-based systems have much to teach others when it comes to protecting the validity of the system and the reliability of assessments made within it, at the same time as protecting the rights of individuals' (p.17). He argues that in Australia, comparability of school-based programmes depends on a combination of devolution of responsibility for assessment design and conduct to teachers who are given appropriate training, plus moderation provided for by common syllabuses and a combination of in-school, district-wide and system-wide meetings to ensure common standards and understanding. By comparison, Withers suggests, statistical moderation is arbitrary and not professionally supportive.

Australian experience thus provides a number of paradoxes. At the forefront of school-based moderation, it is also at the mercy of the arbitrary tertiary entrance score. The growing public confidence in teachers as assessors is squarely matched by reactionary forces within higher education. It is ironic that though these latter institutions have traditionally had no system of external moderation equivalent to the British system of external examining, they represent the major source of resistance to the implementation of more enlightened forms of quality assurance and control in schools.

Overview

Quality assurance procedures which are concerned with the quality of the process of assessment and quality control which is primarily concerned with the products of assessment both need to be judged in terms of their contribution to reliability and validity. Perhaps even more important however, is the issue of utility and how far the assessment procedures being used contribute to, or inhibit the achievement of national educational goals. Although over-polarised, Figures 1 and 2 illustrate the considerable variety that exists in each of the systems mentioned across different courses and levels of education. Given the major changes currently taking place in some systems, it summarises some of the important differences that have been described.

Insert Figures 1 and 2 here.

The origin of moderation procedures was the concern that assessment data should be meaningful; that like any other language, the code of grades or marks should represent an equivalent reality in terms of the standard achieved whether it be between different subjects, different students or different institutions. Moderation is thus closely bound both with notions of utility and of fairness for individuals and for institutions. However, this international review has revealed many tensions in the achievement of such goals. It is apparent that in some countries at least the powerful voice of tradition and associated assumptions and the all important need for public credibility are together inhibiting the development of quality assurance and control procedures more suited to the achievement of the system's goals. Thus norm-referenced scaling procedures designed to rank students for selection continue to be imposed on systems which seek to develop more inclusive, criterion-referenced qualifications. The continued high-status of external written examinations inhibits the effective pursuit of skills and competencies not

amenable to such assessment. Qualifications which do involve more suitable forms of assessment continue to be regarded as of lower status. The valuable personal and professional development which can accrue for both teachers and students from being involved in making assessments (Sadler 1989) is sacrificed to the apparently greater rigour of externally managed assessment.

But as one recent English research project into moderation has found (Shaw and Radnor, 1993), despite being promised no pay-off for school-based assessment and moderation except the prospect of a greater investment of time, resources and effort, teachers' commitment to their students leads them to favour student-centred rather than bureaucratic assessment. In particular in this project they favoured

> 'consensus moderation over inspection and statistical moderation because this implies that standards are constructed in social interaction, not drawn down to the humdrum reality of the moderation exercise from some ideal Platonic realm; and because the social negotiation model really does offer some pay-off for teachers in the form of participation, empowerment and professional development. It is likely to improve teaching and learning by developing awareness, raising self-respect and providing feedback via the teachers to the students ... (standards) are thus socially negotiated, though in the real world where people and groups have different power ratification by representatives of society wider than those present at the moderation meetings is needed for acceptability to the users' (p.249)

This quotation summarising the findings of a small regional certification initiative also summarises the findings of this international review. On the one hand it raises the key issue of credibility and the need, particularly in 'high-stakes' environments where much is at stake for the individual and/or the institution, for trustworthiness and transparency of the data in the interests of both efficiency and social justice. On the other, it raises the issue of professional development and the contribution of quality assurance and control arrangements to both individual and institutional development through appropriate feedback. Clearly differences in these two respects are likely to influence the complexity in character of the moderation methods used. National culture and tradition is likely to provide a further source of influence as it reflects public acceptability of different approaches. As Figure 3 sets out, the choice between more bureaucratic or more professional approaches to quality assurance and control, represents a 'trade-off' between different costs and benefits.

Insert Figure 3 here

The lesson from this review however is that for such decisions to be made <u>rationally</u> rather than arbitrarily, there is an urgent need

1. for international research evidence concerning the relative effectiveness of the full range of moderation approaches in terms of (a) consistency of standards, (b) professional development, (c) public credibility;

2. for policy-makers to be challenged to identify their assessment priorities in these terms and to demonstrate how well the moderation procedures being employed meet these,

3. to explore in the light of national educational culture and assessment traditions, the most effective strategies for bringing about change in both professional practice and public attitudes.

I am greatly indebted to Dr. Caroline Gipps and Professor Margaret Sutherland for some of the material in this paper.

REFERENCES

BAKER, E.L., O'NEIL, H.F. LINN, R.L. (1991) 'Policy and Validity Prospects for Performance-Based Assessment'. Paper presented at the annual meeting of the American Psychological Association, August .

BERLAK, H., NEWMANN, F.M., ADAMS, E., ARCHBALD, D.A., BURGESS, T., RAVEN, J. & ROMBERG, T.A. (eds) *Towards a New Science of Educational Testing and Assessment.* Albany: State University of New York

BAUMGART, N. & McBRYDES, B. (1992) 'Admission to Higher Education: Contrasts in Two Neighbouring Australian States'. Paper presented to 18th Annual Conference of the IAEA, Dublin.

CAROFF, A. (1991) Evaluation et validation des acquis: le diplôme national du brevet, in: *Ministère de L'Education, Rapport sur l'Envaluation des acquis des Elèves à la fin des cycles d'Apprentissage.* Ministry of Education: Paris

CHISHOLM, L. (1987) Vorsprung ex machina? Aspects of curriculum and assessment in cultural comparison. *British Journal of Education Policy*, Vol.2, no.2, pp 149-59

CROSSLEY, M. & BROADFOOT, P. (1992) 'Comparative and International Research in Education: scope, problems and potential.' *British Educational Research Journal* Vol.18 No.2

JAMES, M. & CONNOR, C. (1993) 'Are Reliability and Validity Achievable in NC Assessment: Some observations at KS1 in 1992?' *Curriculum Journal* Vol.4, No.1

FALLON, M. (1993) Article in TES 16.4.93

KREEFT, Henk P.J. (1990) *Examinations in Europe: Three Major Streams*

LINN, R. (1991) 'Technical Considerations in the Proposed Nationwide Assessment System for the National Education Goals Panel'. Prepared for the Technical Planning Sub Group of the Goal 3 Resource Group to the National Education Educational Goals Panel

MADAUS, G.E & KELLAGHAN, T. (1991) 'Student Examination Systems in the EC: Lessons for the United States'. Washington DC Office of Technology Assessment

MADAUS, G (1993) 'A National Testing System: Manna from above? An historical/technological perspective.' *Educational Assessment* 1(i) pp 9-26

NEATHER, E.J. (1993) 'The Arbitur Examination.' *Language Learning Journal* No.7 March pp 19-21

NELSON, D. (1993) Personal communication

NEW ZEALAND QUALIFICATIONS AUTHORITY (1992). *Designing a Moderation System*

NEW ZEALAND QUALIFICATIONS AUTHORITY (1992). *Moderation of Assessment: An Introduction for National Standards Bodies*

OFFICE OF TECHNOLOGY ASSESSMENT, US Congress 1992. *Asking the Right Question: Testing in America.* Washington DC, Government Printing Office.

PHILLIPS, D. (1989) 'Crossnational attractions in Education.' *Comparative Education* 25(3)

SADLER, R. (1989) 'Formative Assessment and the Design of Instructional Systems'. *Instructional Science* Vol.18 pp 119-144

SHAW, K.E. & RADNOR, H. (1993) 'Moderation in Assessment: the case of MAP-SW' *Research Papers in Education* Vol.8 June pp 227-249

SHEPARD, L. (1991) 'Psychometricians' beliefs about learning.' *Educational Researcher* Vol. 20 pp 2-16

SUTHERLAND (1993) 'Assessment in Germany', in NISBETT, J. (ed) *Assessment and Curriculum Reform.* OECD

TUCKER, M. and RESNICK, L. B. (1990) 'Setting a new standard: toward an examination system for the United States'. Proposal for the New Standards Project, University of Pittsburgh.

VIVIANI, N. (1990) *The Review of Tertiary Entrance in Queensland.* Report submitted to the Minister for Education, Brisbane.

WITHERS, G. (1992) 'Achieving Comparability of School-based Assessments used within Tertiary Admissions Procedures - the Australian Experience'. Paper given at International Association for Educational Assessment, Dublin.

Fig.1

APPROACHES TO QUALITY ASSURANCE

	Defining Criteria	Accreditation of institutions/ centres	Visits of verifiers	Agreement panels
US	*		*	*
Germany		*		*
Sweden	*			
France		*		
New Zealand	*	*	*	
Australia	*			*

Fig.2

APPROACHES TO QUALITY CONTROL

	Reference/ scaling tests	Inspection of samples by post	Inspection of samples by visits	External Examining	Appeals	Panel Review
US		*		*		*
Germany					*	
Sweden	*			*	*	
France				*	*	
New Zealand	*	*		*	*	
Australia	*	*	*	*	*	*

Fig.3

PRIORITIES AND TRENDS IN SIX DIFFERENT COUNTRIES' APPROACH TO PROVIDING FOR QUALITY IN ASSESSMENT